



HAL
open science

Relative efficiency of morphological characters and molecular markers in the establishment of an apricot core collection

L. Krichen, Jean Marc Audergon, N. Trifi-Farah

► **To cite this version:**

L. Krichen, Jean Marc Audergon, N. Trifi-Farah. Relative efficiency of morphological characters and molecular markers in the establishment of an apricot core collection. *Hereditas*, 2012, 149 (5), pp.163 - 172. 10.1111/j.1601-5223.2012.02245.x . hal-02643301

HAL Id: hal-02643301

<https://hal.inrae.fr/hal-02643301>

Submitted on 28 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Relative efficiency of morphological characters and molecular markers in the establishment of an apricot core collection

L. KRICHEN¹, J. M. AUDERGON² and N. TRIFI-FARAH¹

¹Laboratoire de Génétique Moléculaire, Immunologie et Biotechnologie, Faculté des Sciences de Tunis, Université Tunis-El Manar, Campus Universitaire – Tunis-El Manar, Tunisie

²INRA UR1052, Unité de Génétique et Amélioration des Fruits et Légumes, Centre de Recherches d'Avignon - Domaine St Maurice – BP94, Montfavet Cedex, France

Krichen, L., Audergon, J. M. and Trifi-Farah, N. 2012. Relative efficiency of morphological characters and molecular markers in the establishment of an apricot core collection. – *Hereditas* 149: 163–172. Lund, Sweden. eISSN 1601-5223. Received 3 October 2011. Accepted 17 August 2012.

In order to optimize the management of genetic resources, in most cases a representative sample of the germplasm collections needs to be developed. The establishment of a core collection is thus of major importance either to minimize the cost associated with the management of the associated germplasm or to apply analysis onto representative bases.

In order to select a representative core collection among the Tunisian apricot germplasm of 110 accessions large, the Maximization strategy algorithm was used. This algorithm was shown to be the most convenient when using both morphological traits and molecular markers. Three core collections based on morphological characters, molecular markers or the combined data were compared. Our data indicate that both the molecular and the morphological markers have to be considered to obtain a core collection that represents the global diversity of the 110 accessions. Using this method, a subset of 34 selected accessions was found to represent accurately the 110 accessions present in the whole collection (75 to 100% for the morphological characters and 97% of the molecular markers). These results show that the combination of molecular and morphological markers is an efficient way to characterize the apricot core collection and provides an exhaustive coverage for the analyzed diversity on morphological and genetic bases.

Lamia Krichen, Laboratoire de Génétique Moléculaire, Immunologie et Biotechnologie, Faculté des Sciences de Tunis, Université Tunis-El Manar, Campus Universitaire – Tunis-El Manar, 2092, Tunisie. E-mail: krichlam@yahoo.com

Apricots (*Prunus armeniaca* L.) belong to the family *Rosaceae*, subfamily *Prunoideae*, the *Prunophora* subgenus of the genus *Prunus*. They are found in the five continents; they are adapted to grow from arid to southern temperate climates and are characterized by an extensive variability that is related to their ecological requirements. Nevertheless, apricots are characterized by a high specificity of the existing cultivars (BAILEY and HOUGH 1975; FAUST et al. 1998) with narrow adaptive areas.

Apricots are traditionally cultivated in Tunisia and cultivars of minor economical interest are not used in the intensive modern orchards. Accordingly, a sizable fraction of the autochthonous germplasm was threatened by genetic erosion in particular in the restricted traditional areas of cultivation. Previous studies conducted by VALDEYRON and CROSSA-RAYNAUD (1950), CROSSA-RAYNAUD (1960) and CARRAUT and CROSSA-RAYNAUD (1974) showed that only few of the previously described cultivars have been encountered in the last surveys. In fact, among the 48 traditional cultivars previously described, 26 disappeared according to the recent surveys conducted in the same areas (KRICHEN et al. 2009). Accordingly, the variability described in Tunisia was

threatened as a consequence of intensive agricultural practice which was advantageous for a limited number of cultivars widely propagated while most of the traditional landraces were cultivated in small area and were unknown elsewhere. This erosion of genetic diversity called for an initiative based on genetic resources preservation as soon as possible.

The main goal of germplasm management is to collect and to characterize diverse forms, in particular at the national and regional level (KHADARI et al. 2003). The first criterion to select representative accessions is based on morphological and agronomic traits of interest. Plant breeders routinely use morphological characterization for the initial description and classification of germplasm in order to select valuable genetic resources for direct use by farmers or in breeding programs. However, recent studies show that molecular markers are indispensable for the germplasm management. Consequently, studies on genetic variability of genetic resources under collecting, help to efficiently preserve valuable germplasm and at the same time, avoid the storage of redundant ones which contributes to the germplasm management.

The need to develop collections for efficient conservation and utilization of genetic diversity has led to the development of core collections for many crop plants including tomato (RANC *et al.* 2008), grape (LE CUNFF *et al.* 2008), loquat (MARTÍNEZ-CALVO *et al.* 2008), pearl millet (BHATTACHARJEE and KHAIRWAL 2007), West African yam (MAHALAKSHMI *et al.* 2007), common bean (RODINO *et al.* 2003; LOGOZZO *et al.* 2007), soybean (WANG *et al.* 2006), safflower (DWIVEDI *et al.* 2005), taro (OKPUL *et al.* 2004), groundnut (UPADHYAYA *et al.* 2003), sugarcane (BALAKRISHNAN *et al.* 2000; AMALRAJ *et al.* 2006) and sesame (XIURONG *et al.* 2000).

FRANKEL (1984), BROWN (1989a, 1989b), MARITA *et al.* (2000) and RODINO *et al.* (2003) argue that the purpose of a core is to provide potential end-users with a representative sample of the available genetic variation of the crop gene pool in a subset of a manageable number. The purpose is to improve utilization and accessibility to vast collections of crop germplasm already maintained and characterized by a gene bank. The entire collection must be reduced to a manageable size that can be easily evaluated to generate good data and enhance utilization.

The core collection is defined as a subset of accessions from a larger collection of particular crop plant that captures most of the available genetic diversity of that crop and its wild relatives with a minimum amount of repetitiveness of this germplasm including its geographical variation. This subset must retain the largest part of the diversity (more than 70% of the entire collection diversity) without redundancy and must be small enough to be easily managed. The rest of the collection should be maintained as the reserve collection. The collection can be evaluated extensively and the information could be used to guide a more efficient utilization of the entire collection. The choice of a sampling strategy is critical in the establishment of core collections, in particular when there are several available criteria and methods proposed to build core collections.

FRANKEL (1984) and BROWN (1989a, 1989b) described methods using information on the origin and on the characteristics of the accessions. Before the setting of the core collection, the size of the final collection as well as the degree of genetic similarity or commonality among accessions have to be taken into consideration to then determine groups within the entire collection.

The FRANKEL and BROWN (1984) strategy involves the stratification of the collection and the selection of a representative set by random sampling from each of the classified groups. The accessions are first classified according to the taxonomy (species, subspecies, races) then according to their geographic location (country, state), climate or agro-ecological regions. The clustering within the broad geographic group could be done

using strongly inherited traits. The number of accessions selected from each cluster will depend on the strategy used and the selection of core collection was made after sub-clustering within the identified groups (SPAGNOLETTI-ZEULI and QUALSET 1993; UPADHYAYA *et al.* 2003; MAHALAKSHMI *et al.* 2007). According to XIURONG *et al.* (2000) the hierarchical clustering methodology was chosen after testing and comparing several cluster trees with respect to their balance, expansibility with consideration of the ecological genotype, origin and correlation of identified traits. Ward's method was proven to be the best one for clustering. It is defined by HINTZE (2001) as follow: on the base of the variance minimization within groups; groups are formed so that the pooled within-group sum of squares is minimized. That is, at each step, the two clusters are linked which results in the least increase in the pooled within-group sum of squares (HINTZE 2001). DWIVEDI *et al.* (2005) used the stratified sampling by geographic origin based on Ward's hierarchical clustering and divided the collection into groups or strata and then a simple random sample is selected from each group on the base of passport or characterization data.

BALAKRISHNAN *et al.* (2000) proposed two methods: 1) non-hierarchical cluster analysis with previous identification of clusters number; 2) principal component analysis which consists on the identification of the generalized sum of squares defined as the product of individual numbers and variable numbers constituting the factor space. A comparison of these two methods proved that the most suitable one for core collection identification is the principal component analysis which defines new independent variables, maximizes diversity and avoids redundancy or duplicates.

MARITA *et al.* (2000) developed an algorithm to assist in selecting core collection which maximizes genetic distances among a set of accessions and ranks all other accessions relative to one accession.

GOUESNARD *et al.* (2001) proposed a maximization strategy consisting in the construction of an algorithm for building germplasm core collection by maximizing allelic or phenotypic richness. The methodology to identify the core size depends on the equation of cumulative inertia for successive accessions; the number corresponds to the peak or inflection point of the curve.

DIWAN *et al.* (1995) compared three methodologies for cores selection implying logarithmic method, proportional method and relative diversity method and defined which one of these methods to use depending on conditions and set of data.

Extensive collection in national and international gene banks has been going on for some time. But, as described above, in Tunisia the variability of apricot landraces was severely threatened. The associated risk of

genetic diversity lost induced the initiation of a national core collection policy based on the recollection of the largest genetic diversity in the shortest delays.

For the management of these *ex situ* plant germplasm, three important goals were set. First, all accessions should be characterized in order to eliminate cases of mislabeling and redundancies and to create a complete data base. Second, to keep a minimum of accessions, this should represent a maximum of variability, constituting a core collection. Third, integrate this germplasm in future breeding programs for new cultivars selection.

As for all fruit trees, an *ex situ* collection need to be installed for an optimal management and use of the apricot genetic diversity. Such core will be taken in charge by the national gene bank as far as their evaluation and management is concerned.

In this aim we tried to select the Tunisian apricot germplasm core collection using morphological characters and data on molecular markers. Such core collection is

needed to safeguard all cultivars and particularly the minor ones, to avoid a loss of genetic diversity and to offer an adequate genetic basis of breeding programs that will use the collection as a reference in Tunisia and at a larger scale.

MATERIAL AND METHODS

Plant material

Plants were taken from the principal locations of apricot cultivation in Tunisia. The 14 prospected sites are distributed as follow: north: Ras Jbel and Testour, center: Kairouan and Mahdia, south: Sfax Gabes, Mareth and Jerba, Oases: Gafsa, Tozeur, Nefta, Degache, Tameghza and Midess.

The total collection was based on 110 apricot accessions (Fig. 1) which have been characterized by morphological characters and AFLP molecular markers.

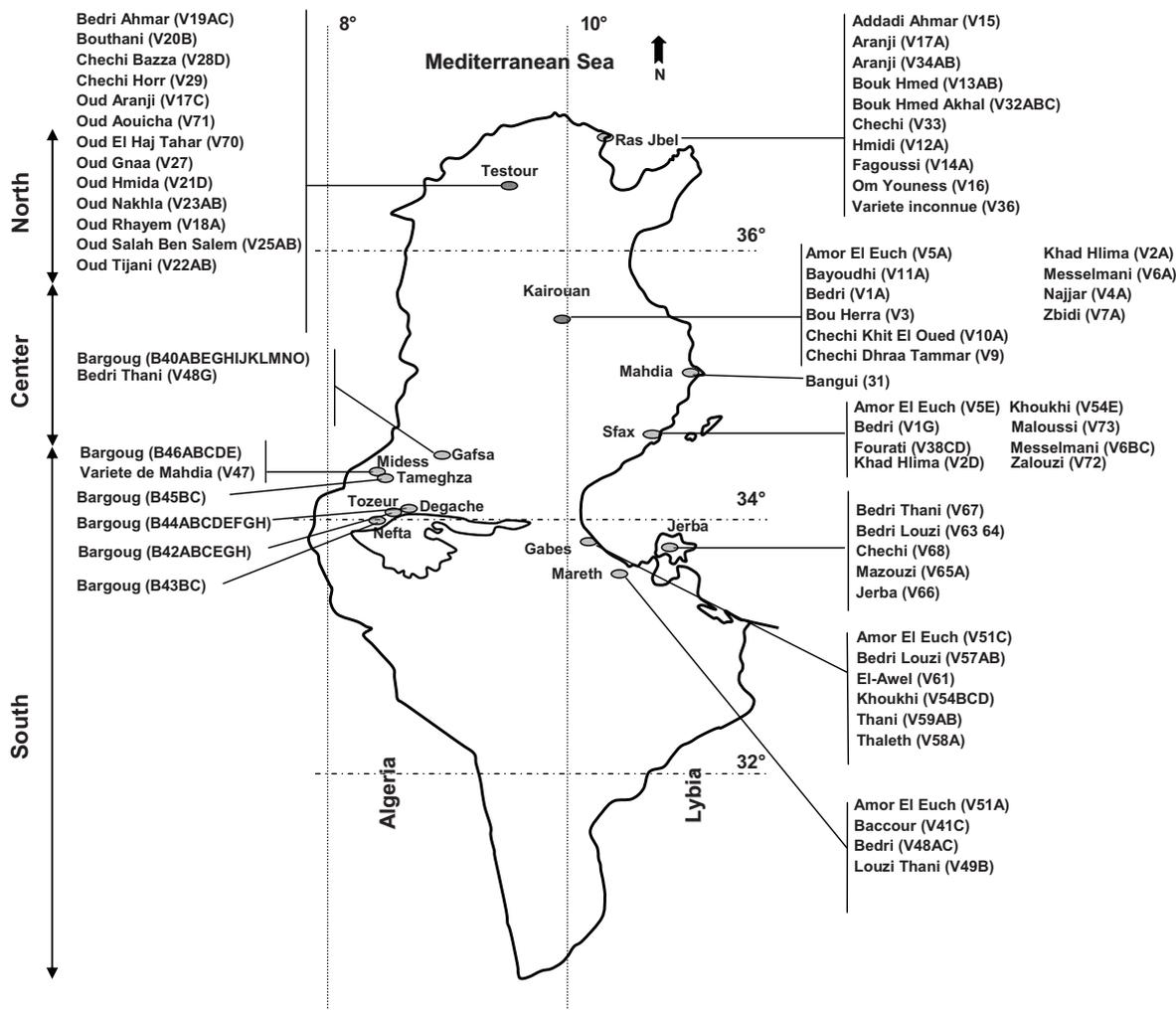


Fig. 1. Geographic origin of the 110 apricot accessions (V: cultivar, B: Bargoug; the number corresponds to the code reference of the accession; alphabetic letters correspond to the repetition of the same nomenclature of the accessions).

Table 1. List of the 34 studied morphological characters in the Tunisian apricot germplasm.

Organ	Variables		
	Quantitative	Qualitative ordinal	Qualitative nominal
Leaves	Leaf blade length (LBL)	Leaf blade: undulation of margin (LBUM)	Leaf blade: shape of base (LBSB)
	Petiole length (LPL)	Petiole anthocyanin coloration of upper side (LPACUS)	Leaf blade: shape of tip (LBST)
	Leaf blade length/leaf blade width (LBL/LBW)	Petiole anthocyanin coloration of lower side (LPACLS)	Leaf blade: angle of tip (LBAT)
	Petiole length/leaf blade length (LPL/LBL)	Intensity of green color of upper side (LIGCLS)	Leaf blade: incisions on margin (LBIM)
Fruits	Fruit weight (FW)	Depth of suture (FDS)	Shape in lateral (or profile) view (FSLV)
	Stone weight (FSW)	Depth of pedicel cavity (FDPC)	Shape in ventral (or frontal) view (FSVV)
	Fruit lateral width/fruit ventral width (FLW/FVW)	Relative area of over color (FROC)	Apex shape (FAS)
	Fruit height/fruit ventral width (FH/FVW)	Firmness of flesh (FFF) Adherence of stone to flesh (FASF) Kernel bitterness (FKB)	Fruit symmetry along the suture (FS) Flesh color (FFC) Ground color of skin (FGC) Fruit surface (FFS) Stone shape (FSS)
Tree		Vigor (TV) Distribution of flower buds (TDFB) One-year old shoot: lenticels number (TLN)	Growth habit (TGH)

Morphological characterization

Morphological characterization have been carried out for apricot fruits, leaves and trees using eight quantitative morphometric traits (measured on a set of 30 fruits and 30 leaves for each accession) and 26 qualitative morphological characters defined in the guidelines for the conduct of tests for distinctness, homogeneity and stability established by the international union for the protection of new varieties of plants (UPOV 1979). These characters are synthetized in Table 1.

Molecular markers

DNA was extracted from young and fresh leaves according to maxi-prep protocol described by BERNATZKY

and TANKSLEY (1986). DNA was digested by EcoRI and MseI endonucleases, and specific amplifications were assessed according to Vos et al. (1995) with a slight modification based on the increase of the concentrations as described by KRICHEN et al. (2008) using five AFLP primer combinations (E32-M36, E33-M40, E35-M35, E39-M42 and E35-M45) (Table 2). EcoRI primers were radioactively labeled using [γ - 33 P] ATP. PCR products were run on denaturing polyacrylamide gel (5%) and exposed to X-ray films for two days.

Statistical analysis and core collection building

Euclidian coefficient for quantitative traits was used to estimate the distance between accessions taken by pair.

Table 2. Polymorphic AFLP markers identified for the five primer combinations.

Primer combination EcoRI/MseI	Respective sequences of the 3 selective nucleotides	Total no. of bands	No. of monomorphic bands	No. of polymorphic bands	Percentage of polymorphism
E 32-M 36	E-AAC/M-ACC	52	9	43	82.69
E 33-M 40	E-AAG/M-AGC	53	18	35	66.04
E 35-M 35	E-ACA/M-ACA	52	18	34	65.38
E39-M42	E-AGA/M-AGT	56	7	49	87.50
E35-M45	E-ACA/M-ATG	60	17	43	71.67
Mean		54.6	13.8	40.8	74.73
Total		273	69	204	

The EcoRI–MseI primer combinations (HAGEN et al. 2002).

The Core collection has been defined using the maximization strategy algorithm implemented in Mstrat software V4.1 (GOUESNARD et al. 2001) and Microcal software Origin V6 (Microcal Software, <www.microcal.com>) for the Mstrat graphics elaboration. Both random (R) and maximization (M) strategies were computed eight times for each data set and compared. Accessions with the highest frequencies allowed the selection of the representative accessions of the core. Optimal core size needs to be represented by the minimum number of accessions with the maximum representativeness of the global variability with a minimum or no redundancy. Redundancy allowed the comparison between maximization and random strategies to assess the optimal number of accessions representing the core collection. The methodology to identify the core size depends on the equation of cumulative inertia for successive accessions; the number corresponds to the peak or inflection point of the curve. The construction of the core collection was performed using 15 replications for 50 maximum iterations with Shannon index evaluation.

Accessions with unique characteristics (unique characteristic is one modality of the morphological character which is observed with only one accession), even though showing an extremely low frequency, were included in the core collection. Subsequently, variation for each trait within the core and base collection was compared to ensure consistency of genetic variability.

RESULTS

Evaluation of diversity and polymorphism in morphological and molecular characters

All 34 quantitative and qualitative morphological traits were polymorphic. The available modalities for the qualitative and quantitative variables were represented by the studied germplasm. They allowed distinguishing each one of the 110 accessions. For molecular markers, AFLP showed an average total band number (273), an average number of polymorphic bands (204), and an average percentage of polymorphism (75%) for each primer combination (Table 2).

Core collection construction

Because of differences in genetic and morphological diversity, core collections were sampled using both genetic and morphological diversity. Based on 34 polymorphic morphological characters and 204 polymorphic molecular markers, maximization strategy algorithm was used to construct Tunisian apricot germplasm core collection using the Mstrat software as described by GOUESNARD et al. (2001).

The construction of a core collection allowed the selection of several core sizes from the global diversity (110 apricot accessions) according to the Mstrat strategy of selection. It could be selected using the visualization of optimum and random means for active and target variables (active variables are those called Markers, 'Target' variable means that Mstrat will compute the score realized on these variables using active variables). For that reason, Random (R) and Maximization (M) strategies were compared. Thus, three different core collections were elaborated; the first one is based on the 34 morphological characters, the second one includes the 204 AFLP polymorphic molecular markers and the third one represents the combined morphological-AFLP data. Eight cores of the selected size were constructed and compared. The same procedure has been considered for the three cores construction.

As for the computation of the redundancy for active and target variables, the results showed that the optimization strategy allow us to reach rapidly the optimal size of the core which corresponds to the beginning of the plateau of the curve. Similar results were reported by GOUESNARD et al. (2001) which indicates that the inflection point of the M curve provides the optimal size for a core collection.

Again, the combined morphological-AFLP data core, the plotting of all optimum (Maximization method) and random values related to mean values (Fig. 2) and all points values (Fig. 3) for active and target variables showed that the plateau was reached more rapidly with the M method than with the R one for active variables (Fig. 2a). The results showed that the ideal size of the core collection obtained at the plateau of the OPT curve is around 20 individuals.

The core collection representative of the global genetic diversity of the studied sample was selected after several core collection constructions using Mstrat software. The accessions representing the core were selected in relation with their high frequency of sampling in the different core collections.

Results permitted the construction of cores allowing the choice of the 23 accessions that are highly iterated. They correspond to:

- Bargougs: B40A, B40K, B40L, B40M, B46D,
- Cultivars: 'Chechi Khit El OuedV10A', 'Bouk HmedV13B', 'Chechi Dhraa TammarV9', 'Chechi HorrV29', 'Amor El EuchV5A', 'Oud AouichaV71', 'Oud TijaniV22B', 'Oud GnaaV27', 'BanguiV31', 'Bouk Hmed AkhalV32B', 'Khad HlimaV2A', 'BaccourV41C', 'Bedri AhmarV19A', 'NajjarV4A', 'Oud Salah Ben SalemV25B', 'Amor El EuchV51C', 'Oud HmidaV21D', 'Oud NakhlaV23A'.

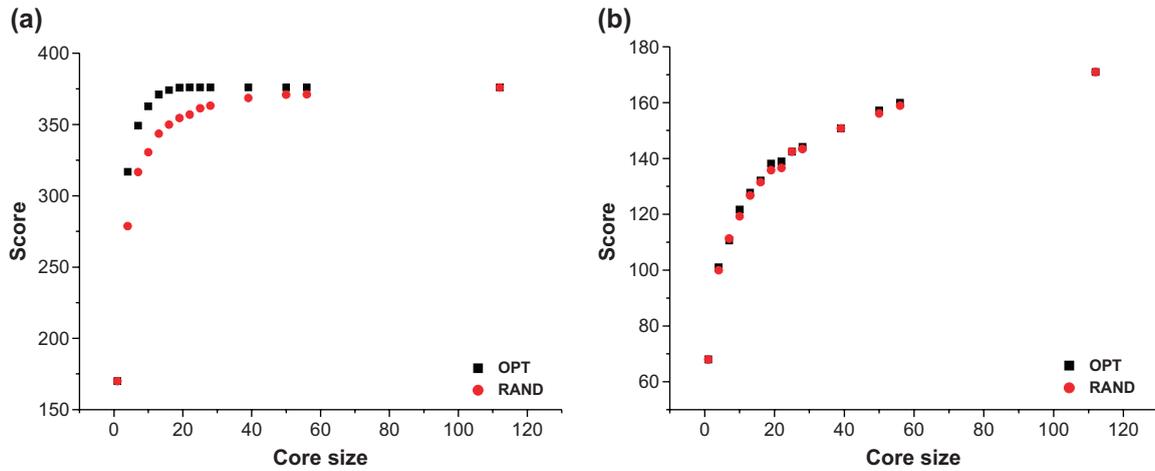


Fig. 2a–b. Plotting of optimization (OPT) (Maximization method) and randomization (RAND) (random method) related to the mean values for active variables (a) and target variables (b).

This core collection needs to be completed with 11 other accessions representing specific molecular markers or rare modalities of morphological characters corresponding to:

- Bargougs: B44C, B44D, B46B, B46E,
- Cultivars: ‘Oud TijaniV22A’, ‘JerbaV66’, ‘Variete deMahdiaV47’, ‘ChechiBazzaV28D’, ‘ChechiV68’, ‘AranjiV17C’, ‘BayoudhiV11A’.

As a result, the apricot core collection was represented by a total of 34 accessions.

If considering the core collection based exclusively on the morphological traits, the plotting of all optimum (Maximization method) and random values related to mean values showed that the plateau was reached more rapidly with maximizing strategy method than with random one for active variables (Fig. 4a) and that the

ideal size of the core collection is about 10 accessions corresponding to:

- Bargougs: B40H, B40K, B40M, B42C,
- Cultivars: ‘Amor El EuchV5A’, ‘Oud AouichaV71’, ‘Chechi HorrV29’, ‘Oud GnaaV27’, ‘Bayoudhi V11A’, ‘Amor El Euch V51C’.

Two accessions, ‘JerbaV66’ and ‘AranjiV17C’, with specific modalities for the variables LPL and LPACLS, need to be added to this core increasing the total number of the accessions to 12.

For the core collection based on the AFLP molecular markers, the plotting of all optimum (Maximization method) and random values related to mean values revealed also that the plateau was reached more rapidly with maximizing strategy method than with random one

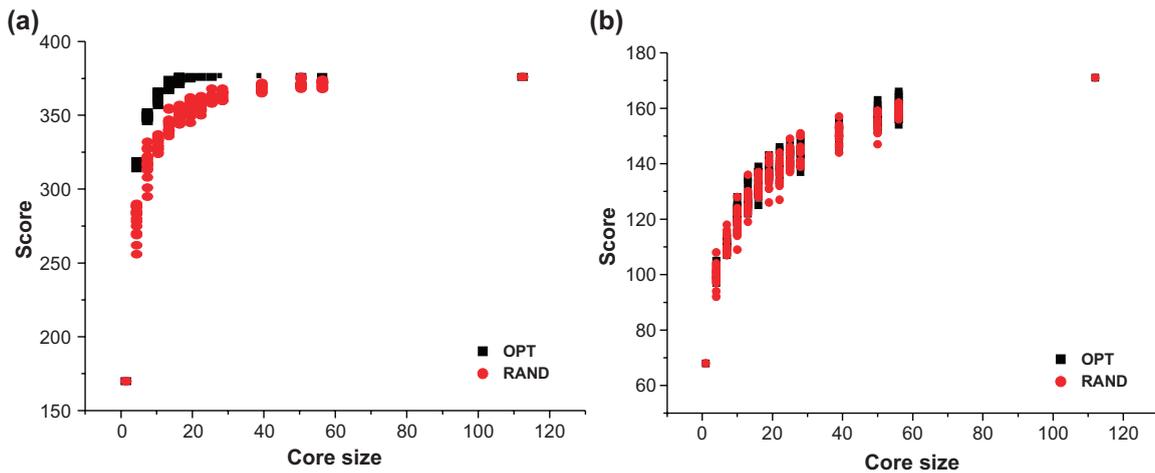


Fig. 3a–b. Plotting of optimization (OPT) (Maximization method) and randomization (RAND) (random method) related to the values for all points relatively to active variables (a) and target variables (b).

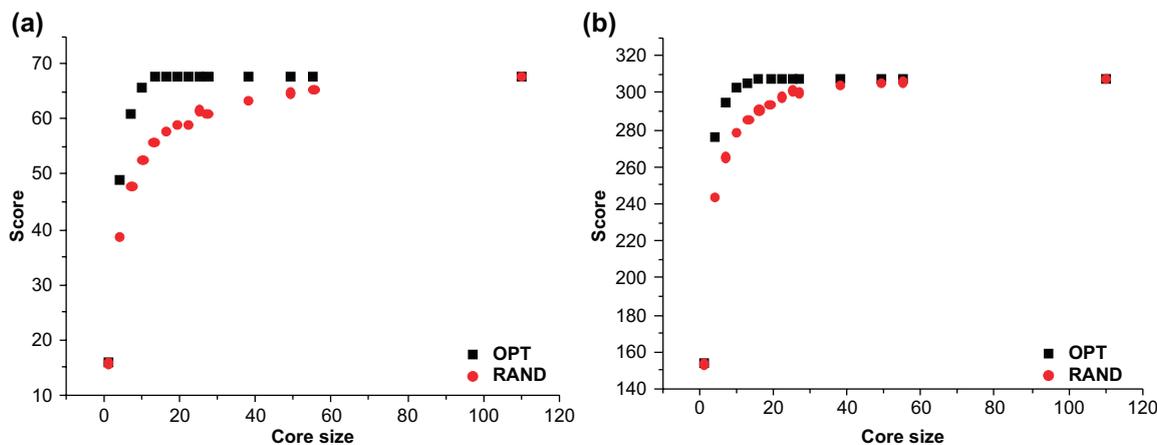


Fig. 4a–b. Plotting of optimization (OPT) (Maximization method) and randomization (RAND) (random method) related to the mean values relatively to active variables related to the morphological data (a) and the AFLP data (b).

for active variables (Fig. 4b). This allowed the identification of the ideal size of the core collection corresponding to 18 individuals listed as follows:

- Bargougs: B40G, B40J, B45B,
- Cultivars: ‘Amor El EuchV5A’, ‘Chechi Khit El OuedV10A’, ‘BayoudhiV11A’, ‘Bouk HmedV13B’, ‘AranjiV17A’, ‘Bedri V1G’, ‘Bedri AhmarV19A’, ‘Oud TijaniV22B’, ‘Chechi Dhraa TammarV9’, ‘Khad HlimaV2A’, ‘BaccourV41C’, ‘Amor El Euch V51C’, ‘Bouk Hmed AkhalV32B’, ‘NajjarV4A’, ‘Bangui V31’.

The additive list of accessions representing rare markers is composed by nine accessions:

- Bargougs: B44C, B44D, B46B, B46D, B46E,
- Cultivars: ‘Oud TijaniV22A’, ‘Chechi BazzaV28D’, ‘AranjiV17C’, ‘Variete de MahdiaV47’.

Consequently, this final core size reached 27 accessions.

Core collection validation and comparison

Richness of a collection of accessions for such a qualitative variable was defined as the number of classes represented among the accessions (GOUESNARD et al. 2001).

Comparison between morphological characters variability observed for the entire collection (110 accessions) and the morphological variability of each of the selected core collection is shown in Table 3.

The intervals of variance were compared for quantitative variables and modalities observed were compared for the qualitative variables. Results showed that for the quantitative variables FW, FSW, LBL, LPL, LBL/LBW and LPL/LBL, the variability of the combined data core collection (34 accessions) corresponds respectively to 98%, 80%, 75%, 72%, 71% and 88% of the variability of the 110 accessions. For qualitative characters, 100%

of the variability of the 110 accessions was represented by the 34 accessions of the core collection (Table 3).

Among the 204 polymorphic AFLP markers, only seven markers were not represented by the combined data core collection, thus, the 34 accessions covers 97% of the genetic variability of the 110 accessions.

Differences between the morphological variability, the molecular diversity of the entire collection (110 accessions) and the subset of the core collection (34 accessions) were found to be non-significant for all the morphological and molecular markers recorded indicating that the core of 34 accessions is well representative of the global diversity.

The combination of the morphological variability and the molecular diversity shows that the core of 34 accessions represents from 70 to 100% of the existing variability (110 accessions). Accordingly, the combination of morphological and molecular markers is an efficient tool for characterizing the apricot core collection and will be valid to distinguish other accessions which can be introduced into the collection with more than 70% of the entire collection diversity.

The elaborated core collection by the morphological characters showed that all the modalities of the qualitative variables are represented by the core set of 11 accessions at a level of 100% except for the characters TDFB (33%), FKB (75%), TGH (75%), LBAT (67%), LBIM (75%), FFS (50%); while for the quantitative traits; the representativeness is about 94%, 54%, 52%, 67%, 85%, 64%, 82% for FW, FSW, LBL, LPL, FLW/FVW, LBL/LBW and LPL/LBL, respectively (Table 3). Noteworthy that non-significant difference was observed for the qualitative characters even when values are less than 70% of the global variability. The set of 12 accessions is less representative of the global variability if we refer to these percentages of the interval of variance of each trait.

Table 3. Comparison of the morphological characters variability between the entire collection and the different core size collections (interval of variance for quantitative variables, observed modalities for the qualitative variables).

Morphological variables	Global collection (110 accessions)	Morphological core collection (12 accessions)	Combined data core collection (34 accessions)
FW	2.71 to 54.31	5.75 to 54.31 (94%)	3.45 to 54.31 (98.6%)
FSW	0.70 to 4.6	1.13 to 3.23 (54%)	0.75 to 3.88 (80%)
LBL	4.28 to 9.33	5.51 to 8.12 (52%)	5.21 to 9.01 (75%)
LPL	1.69 to 4.78	1.93 to 4.01 (67%)	2.19 to 4.42 (72%)
FLW/FVW	0.99 to 1.38	1.05 to 1.38 (85%)	0.99 to 1.38 (100%)
FH/FVW	0.92 to 1.73	0.92 to 1.73 (100%)	0.92 to 1.73 (100%)
LBL/LBW	0.86 to 1.41	0.86 to 1.21 (64%)	0.86 to 1.25 (71%)
LPL/LBL	0.30 to 0.63	0.32 to 0.59 (82%)	0.30 to 0.59 (88%)
TV	1,2,3	1,2,3 (100%)	1,2,3 (100%)
TDFB	1,2,3	2 (33%)	1,2,3 (100%)
TLN	1,2,3	1,2,3 (100%)	1,2,3 (100%)
LBUM	1,2,3	1,2,3 (100%)	1,2,3 (100%)
LPACUS	1,2,3	1,2,3 (100%)	1,2,3 (100%)
LPACLS	1,2,3,4,5	1,2,3,4,5 (100%)	1,2,3,4,5 (100%)
FDS	1,2,3	1,2,3 (100%)	1,2,3 (100%)
FDPC	1,2,3	1,2,3 (100%)	1,2,3 (100%)
FROC	1,2,3,4	1,2,3,4 (100%)	1,2,3,4 (100%)
FFF	1,2,3	1,2,3 (100%)	1,2,3 (100%)
FASF	1,2,3,4	1,2,3,4 (100%)	1,2,3,4 (100%)
FKB	1,2,3,4	2,3,4 (75%)	1,2,3,4 (100%)
LIGCLS	1,2,3	1,2,3 (100%)	1,2,3 (100%)
TGH	2,3,4,5	2,4,5 (75%)	2,3,4,5 (100%)
LBSB	1,2,3,4	1,2,3,4 (100%)	1,2,3,4 (100%)
LBST	1,2,3	1,2,3 (100%)	1,2,3 (100%)
FSLV	1,2,3,4	1,2,3,4 (100%)	1,2,3,4 (100%)
FSVV	1,2,3	1,2,3 (100%)	1,2,3 (100%)
LBAT	1,2,3	2,3 (67%)	1,2,3 (100%)
LBIM	1,2,3,4	1,3,4 (75%)	1,2,3,4 (100%)
FAS	1,2,3,4	1,2,3,4 (100%)	1,2,3,4 (100%)
FS	1,2	1,2 (100%)	1,2 (100%)
FFS	1,2	1 (50%)	1,2 (100%)
FGC	1,2,3,4,5	1,2,3,4,5 (100%)	1,2,3,4,5 (100%)
FFC	1,2,3,4,5	1,2,3,4,5 (100%)	1,2,3,4,5 (100%)
FSS	1,2,3	1,2,3 (100%)	1,2,3 (100%)

When considering the AFLP core collection, we conclude that the 27 accessions enclosed 96% of the global genetic diversity and that only 9 markers were not represented by the core set. This difference is not significant showing that the core set of 27 accessions is an accurate representation of the molecular diversity of the 110 accessions.

The comparison of the three core collections showed that the core issued from the combination of the two cores from morphological and AFLP markers is the most efficient.

On the other hand, since Ward's minimum variance hierarchical clustering dendrogram was considered by XIURONG et al. (2000) as the most suitable for core selection when constructed with the same data base, we constructed a Ward's dendrogram. This resulted in the

selection of a representative collection with 39 accessions among the 110 studied. Comparison of the two cores selected on the basis of morphological and molecular markers, using the Mstrat and Ward's methods shows a strong link between the two selected cores with almost 60% of similarity (results not shown).

DISCUSSION

Several methods and strategies for constructing core collections have been proposed and compared covering the randomization, the maximization, the logarithmic, the passport data etc. (DIWAN et al. 1995, XIURONG et al. 2000, DWIVEDI et al. 2005). As suggested by GOUESNARD et al. (2001) and HAOUANE et al. (2011), the maximization

strategy seems to be the most suitable for core collection construction among unknown germplasm variation.

The two methods used are: 1) the stratified method based on genetic distances and clustering. 2) The maximization vs randomization strategy which maximizes the variability as proposed by GOUESNARD et al. (2001).

The comparison showed that the method capturing the maximum of genetic diversity is the maximization strategy algorithm in Mstrat which is always superior to the random strategy as shown by GOUESNARD et al. (2001) and HAOUANE et al. (2011).

The difference between the (R) and the (M) strategy curves suggested that the M method was better for core collection establishment. For sampling core collection, the gain when scoring with the maximization strategy was higher than with random strategy as shown by RANC et al. (2008). This is not surprising because M strategy examines all possible core collections and singles out those that maximize the number of observed alleles at the marker loci. These could be then chosen as final candidates for the core. The expected superiority of this marker-based method took into consideration the correlation between observed allelic richness at the marker loci and allelic richness on other loci (GOUESNARD et al. 2001). It was also evident that the M strategy was well adapted when the accessions came from populations with restricted gene flow which is the case of apricot. Accordingly, the Mstrat software allowed to interactively explore the consequences of the sampling procedure for different sets of traits and to define a robust minimum size for the core collection (GOUESNARD et al. 2001).

For the first time in Tunisia, a national core collection for apricot material has thus been selected. The construction of three cores based on morphological traits, AFLP molecular markers and the combination of the morphological and the molecular data demonstrated that the use of both morphological and molecular markers is very efficient in cores construction because of the complementarity of information they bring. Thus, the core that represents best the global diversity of the 110 accessions is based on the combined data and will be privilege.

The comparison between the core collection and the global collection diversity demonstrated that our core collection is highly representative of the studied germplasm. In addition to the optimal size of the core defined as 20 accessions and selected by the core construction, accessions representing the rare alleles or modalities of the morphological characters with low frequencies were added to the selected core. HAOUANE et al. (2011) have come to a similar conclusion. The selected core captures more than 70% of the global diversity.

The comparison between the Mstrat and Ward's methods showed an approximate similarity in the selected accessions for the core building. In addition,

the 110 accessions are representative of the different geographical regions for apricot culture in Tunisia.

Morphological description completed with molecular marker diversity gives optimal representativeness of the germplasm diversity in a core set without redundancy. This is an efficient tool for genotyping the apricot core collection and will distinguish other accessions which can be introduced into the collection.

Conclusion

The establishment of a core collection will be an indispensable aid in conserving apricot diversity in Tunisia.

As a consequence of this work, all of the 110 studied accessions need to be maintained in an ex-situ collection and the core that captures more than 70% of the diversity, will be useful to develop new breeding programs, association genetics studies and to select new cultivars. The core collection can be evaluated extensively and the information derived could be used to guide more efficient utilization of the entire collection. It should be revised periodically to take into account any additional accessions and information when available.

This core collection should simplify management and enhance the use of apricot genetic resources. It will also be an important entry point to further research and the exploitation of the genetic resources available in Tunisian apricots.

This wide variability could be utilized for further improvement of the crop in enhancing the genetic potential for yield and also in alleviating biotic and abiotic stress factors. Accordingly, the genetically diverse germplasm based on phenotypic and molecular diversity will be made available to breeders to enhance the genetic potential of apricot crop. The core would also provide a guideline to the curator while acquiring new accessions in the gene bank collection.

Such work could enable the enlargement of the core at the Mediterranean basin scale and the establishing of a large core collection representative of the species and the countries.

Acknowledgements – This work was supported by a grant from the French–Tunisian Cooperation - Comité Mixte pour la Coopération Universitaire (CMCU) Project and the Tunisian Ministry of Higher Education and Scientific Research. Authors would like to thank Pauline Audergon for the English revision of the manuscript.

REFERENCES

- Amalraj, A. V., Balakrishnan, R., Jebadhasn, A. W. et al. 2006. Constituting a core collection of *Saccharum spontaneum* L. and comparison of three stratified random sampling procedures. – Genet. Resour. Crop Evol. 53: 1563–1572.

- Bailey, C. H. and Hough, L. F. 1975. Apricots. – In: Janick, J. and Moore, J. N. (eds), *Advances in fruit breeding*. Purdue Univ. Press, pp. 367–383.
- Balakrishnan, R., Nair, N. V. and Sreenivasan, T. V. 2000. A method for establishing a core collection of *Saccharum officinarum* L. germplasm based on quantitative-morphological data. – *Genet. Resour. Crop Evol.* 47: 1–9.
- Bernatzky, R. and Tanksley, S. D. 1986. Genetics of actin-related sequences in tomato. – *Theor. Appl. Genet.* 72: 314–321.
- Bhattacharjee, R. and Khairwal, I. S. 2007. Establishment of a pearl millet [*Pennisetum glaucum* (L.) R. Br.] core collection based on geographical distribution and quantitative traits. – *Euphytica* 155: 35–45.
- Brown, A. H. D. 1989a. Core collections: a practical approach to genetic resources management. – *Genome* 31: 818–824.
- Brown, A. H. D. 1989b. The case for core collections. Core collections: a practical approach to genetic resources management. – In: Brown, A. H. D., Frankel, O. H., Marshall, D. R. et al. (eds), *The use of plant genetic resources*. Cambridge Univ. Press, pp. 136–155.
- Carraut, A. and Crossa-Raynaud, P. 1974. L'amélioration des variétés d'abricotier en Tunisie. – *Ann. Natl Inst. Recherche Agron. Tunisia* 47: 33.
- Crossa-Raynaud, P. 1960. Problèmes d'arboriculture fruitière en Tunisie. – *Ann. Natl Inst. Recherche Agron. Tunisia* 33: 39–63.
- Diwan, N., McIntosh, M. S. and Bauchan, G. R. 1995. Methods of developing a core collection of annual *Medicago* species. – *Theor. Appl. Genet.* 90: 755–761.
- Dwivedi, S. L., Upadhyaya, H. D. and Hegde, D. M. 2005. Development of core collection using geographic information and morphological descriptors in safflower (*Carthamus tinctorius* L.) germplasm. – *Genet. Resour. Crop Evol.* 52: 821–830.
- Faust, M., Surányi, D. and Nyujtó, F. 1998. Origin and dissemination of apricot. – *Hort. Rev.* 22: 225–266.
- Frankel, O. H. 1984. Genetic perspectives of germplasm conservation. – In: Arber, W., Llimensee, K., Peacock, W. J. et al. (eds), *Genetic manipulations: impact on man and society*. Cambridge Univ. Press, pp. 161–170.
- Frankel, O. H. and Brown, A. H. D. 1984. Current plant genetic resources – a critical appraisal. – In: Chopra, V. L., Joshi, B. C., Sharma, R. P. et al. (eds), *Genetics: new frontiers*, vol. IV. Oxford and IBH Publishing Company, New Delhi, India, pp. 1–63.
- Gouesnard, B., Bataillon, T. M., Decoux, G. et al. 2001. MSTRAT: an algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. – *J. Hered.* 92: 93–94.
- Hagen, L. S., Khadari, B., Lambert, P. et al. 2002. Genetic diversity in apricot revealed by AFLP markers: species and cultivars comparisons. – *Theor. Appl. Genet.* 105: 298–305.
- Haoouane, H., El Bakkali, A., Moukhli, A. et al. 2011. Genetic structure and core collection of the World Olive Germplasm Bank of Marrakech: towards the optimised management and use of Mediterranean olive genetic resources. – *Genetica* 139: 1083–1094.
- Hintze, J. 2001. NCSS and PASS. Number Cruncher Statistical System, Kaysville, Utah, USA. <www.ncss.com>
- Khadari, B., Breton, C., Moutier, N. et al. 2003. The use of molecular markers for germplasm management in a French olive collection. – *Theor. Appl. Genet.* 106: 521–529.
- Krichen, L., Martins, J. M. S., Lambert, P. et al. 2008. Using AFLP markers for the analysis of the genetic diversity of apricots in Tunisia. – *J. Am. Soc. Hort. Sci.* 133: 204–212.
- Krichen, L., Trifi-Farah, N., Marrakchi, M. et al. 2009. Evaluation of the current apricot variability in Tunisia – comparison with previously described cultivars. – *Acta Hort.* 814: 113–120.
- Le Cunff, L., Fournier-Level, F., Laucou, V. et al. 2008. Construction of nested genetic core collection to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp. *sativa*. – *BMC Plant Biol.* 8: 31. doi: 10.1186/1471-2229-8-31
- Logozzo, Ga., Donnoli, R., Macaluso, L. et al. 2007. Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. – *Genet. Resour. Crop Evol.* 54: 1763–1779.
- Mahalakshmi, V., Atalobhor, Q. Ng. J., Ogunsoola, D. et al. 2007. Development of a West African yam *Discorea* spp. core collection. – *Genet. Resour. Crop Evol.* 54: 1817–1825.
- Marita, J. M., Rodriguez, J. M. and Nienhuis, J. 2000. Development of an algorithm identifying maximally diverse core collections. – *Genet. Resour. Crop Evol.* 47: 515–526.
- Martínez-Calvo, J., Delia-Gisbert, A. M., Alamar, C. et al. 2008. Study of a germplasm collection of loquat (*Eriobotrya japonica* Lindl.) by multivariate analysis. – *Genet. Resour. Crop Evol.* 55: 695–703.
- Okpul, T., Singh, D., Gunua, T. et al. 2004. Assessment of diversity using agro-morphological traits for selecting a core sample of Papua New Guinea taro (*Colocasia esculenta* (L.) Schott) collection. – *Genet. Resour. Crop Evol.* 51: 671–678.
- Ranc, N., Muñoz, S., Santoni, S. et al. 2008. A clarified position for *Solanum lycopersicum* var. *cerasiforme* in tomato (*Solanaceae*) evolutionary history. – *BMC Plant Biol.* 8: 130: 16p. doi: 10.1186/1471-2229-8-130
- Rodino, A. P., Santalla, M., De Ron, A. M. et al. 2003. A core collection of common bean from the Iberian Peninsula. – *Euphytica* 131: 165–175.
- Spagnoletti-Zeuli, P. L. and Qualset, C. O. 1993. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. – *Theor. Appl. Genet.* 87: 295–304.
- Upadhyaya, H. D., Ortiz, R., Bramel, P. J. et al. 2003. Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. – *Genet. Resour. Crop Evol.* 50: 139–148.
- UPOV 1979. Principes directeurs pour la conduite de l'examen des caractères distinctifs de l'homogénéité et de la stabilité: abricotier (*Prunus armeniaca* L.). Release in 2005. <www.upov.int/restrict/fr/index.html>
- Valdeyron, G. and Crossa-Raynaud, P. 1950. Les fruits de Tunisie. – *Ann. Natl Inst. Recherche Agron. Tunisia* 23: 65–82.
- Vos, P., Hogers, R., Bleeker, M. et al. 1995. AFLP: a new technique for DNA fingerprinting. – *Nucleic Acids Res.* 23: 4407–4414.
- Wang, L., Guan, Y., Guan, R. et al. 2006. Establishment of Chinese soybean (*Glycine max*) core collection with agronomic traits and SSR markers. – *Euphytica* 151: 215–223.
- Xiurong, Z., Yihzhong, Z., Yang, C. et al. 2000. Establishment of sesame germplasm core collection in China. – *Genet. Resour. Crop Evol.* 47: 273–279.